(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 09065287 A

(43) Date of publication of application: 07 . 03 . 97

(51) Int. Cl H04N 5/93
G06T 13/00
H04N 5/268

(21) Application number: 07210409

(22) Date of filing: 18 . 08 . 95

(71) Applicant: HITACHI LTD

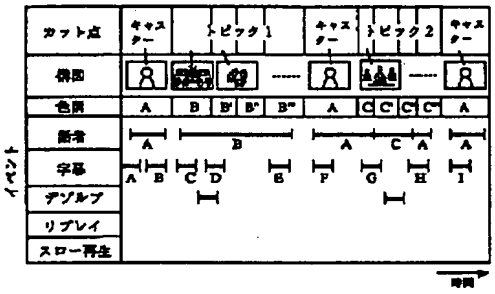
(72) Inventor: NAGASAKA AKIO
MIYATAKE TAKAFUMI
FUJITA TAKEHIRO
TANIGUCHI KATSUMI

(54) METHOD AND DEVICE FOR DETECTING CHARACTERISTIC SCENE FOR DYNAMIC IMAGE

(57) Abstract:

PROBLEM TO BE SOLVED: To obtain information assisting the user for video image selection by discriminating whether or not a scene is an important scene in a video image, specifying its range simply at a high speed, discriminating and classifying to which field a video image belongs (news, sport relay or the like).

SOLUTION: This device is provided with a means entering an object dynamic image to a processer in time series in the unit of frames, a means buffering plural frames entered in the past in the processing unit, and a means discriminating whether or not a feature variable of the buffered frame has a characteristic in which the feature variable approaches that of the newest frame monotonously in the order from the older frames. Furthermore, when it is discriminated true by the discrimination means, a means is provided to extract a video period till a succeeding special video effect is detected or a prescribed after the frame as an important scene. Moreover, in addition to the special video effect, a means detecting a change and a state of various video images (change of cut, display of caption, layout or the like) and a means discriminating kinds of video images based on the appearing order and combination is provided.



JP No. 9-65287

(57) [Abstract]

[Object] To perform judgement whether or not a scene is an important scene in a video image, and specifying its range simply at a high speed. Besides, to judge to which field (news, sport relay, or the like) a video image belongs and classify it to provide information assisting the user to select the video image.

[Constitution] There are provided means for inputting a dynamic image as an object to a processing unit in time series in frame units, means for buffering plural frames inputted to the processing unit in the past, and means for judging whether or not a feature variable of the buffered frame has a characteristic in which the feature variable approaches that of the newest frame monotonously in the order from the older frames. Furthermore, when judgement made by the judgement means is true, it is judged that there is a special video effect, and there is provided means for extracting, as an important scene, a video period till a definite time elapses after the frame or till a succeeding special video effect is detected. Moreover, means for detecting a change and a state of various video images (change of cut, display of caption, layout or the like) in addition to the special video effect is also provided, and means for discriminating kinds of video images based on the appearing order and combination of those is provided.

[0011] FIG. 1 is an example of a schematic block diagram of a system structure for realizing the present invention. Reference numeral 1 designates a display device such as a CRT, which displays an output screen of a computer 4. Instructions to the computer 4 can be made by using an input device 5 such as a keyboard or a pointing device. A dynamic image reproduction device 10 is a tuner device for receiving broadcast programs of ground wave broadcasting, satellite broadcasting, cable television or the like, or a device for reproducing dynamic images recorded on an optical disk, a video tape or the like. A video signal outputted from the dynamic image reproduction device is sequentially converted into digital image data by an A/D converter 3 and is sent to the computer. In the inside of the computer, the digital image data is inputted to a memory 9 through an interface 8, and is processed by a CPU 7 in accordance with a program stored in the memory 9. In the case where a number (frame number) is allocated in sequence from the head of the dynamic image to each frame of the dynamic image processed by the device 10, when the frame number is sent to the dynamic image reproduction device through a control line 2, the dynamic image of the scene can be called and reproduced. Besides, according to the necessity of processing, various information can be stored in an external information storage device 6. Various data prepared by processes explained below are stored in the memory 9, and are consulted

as the need arises.

[0012] Hereinafter, a method of detecting dissolve as one of a cut change by a special video effect at the selection of an important scene, will be described in detail.

[0013] FIG. 2 shows an example of a flowchart of a dissolve detection program of dynamic images executed on the system shown in FIG. 1. The program is stored in the memory 9, and the CPU 7 sets various variables necessary for the execution of the program to initial values as an initializing process (200). Next, 0 is substituted in the respective elements of m two-dimensional arrays B(x, y) containing brightness values of respective pixels of a past frame image (202). When the size of the frame image is w × h, x takes a value from 0 to w-1, and y takes a value from 0 to h-1. In a process 204, a frame image outputted from the dynamic image reproduction device 10 is taken in (204). In a process 206, a variable eval in which an evaluation value is put is made 0, and an initial value 0 is substituted in a loop counter. Then, following processes 208 to 228 are carried out for all pixels in the frame image.

[0014] In the processes 208 to 228, detection of properties peculiar to the dissolve is carried out. Here, the dissolve is a cut change having a period in which as shown in FIG. 3, frame images A and C of cuts are mixed before and after the change of cuts, like B. A mixture ratio of A and C in B is inverted with a time from a state where A is 100% and C is 0 % at the time of start of the dissolve, and the dissolve is completed at a point of time when A finally becomes 0%

and C becomes 100%. In the case of a light and shade image, when a brightness value of A is Ba, a brightness value of B is Bb, a brightness value of C is Bc, and a mixture ratio of C is α ($0 \leq \alpha \leq 1$), approximation can be made by an expression $Bb = Ba \times (1 - \alpha) + Bc \times \alpha$. When this expression is modified, $Bb = (Bc - Ba) \times \alpha + Ba$ is obtained. In the case of the dissolve where the mixture ratio α is monotonously increased from 0, the value of Bb is also monotonously increased or decreased from Ba to Bc. Accordingly, if brightness values of pixels are always stored in a buffer for the past m frames, and it is checked whether the brightness value is monotonously increased or decreased in the period of the m-frame length, detection of the dissolve can be performed. When the value of m is set to about 8 to 15, excellent results are obtained experimentally.

[0015] First, in a process 208, a brightness value of a pixel expressed by a coordinate (x, y) is substituted in an mth array Bm of two-dimensional arrays B storing brightness values of the past frames. Then, 1 is substituted in a loop counter i, and 0 is substituted in a variable num. Next, a brightness value B1 (x, y) stored in the first array is compared with a value of the mth array Bm(x, y) (212), and subsequently, it is compared whether or not a brightness value Bi(x, y) stored in an ith array is larger than a value of a next array Bi+1(x, y) (214, 216). When B1(x, y) is larger than Bm(x, y), in the case where Bi(x, y) is larger than Bi+1(x, y), the value of num is increased by 1. On the contrary, when B1(x, y) is smaller than Bm(x, y), in the

case where Bi(x, y) is smaller than Bi+1(x, y), the value of num is increased by 1 (218). In a subsequent process 220, the value of Bi+1(x, y) is substituted in Bi(x, y), so that the m arrays B are shifted one by one in sequence, and the brightness values of m frames from the newest frame are always stored as a buffer. In a process 222, the loop counter i is increased by 1, and until i becomes larger than m, the processes are repeated in such a manner that when B1(x, y) is larger than Bm(x, y) at the point of time of the process 212, the procedure returns to the process 214, if not, it returns to the process 216 (224). When the variable num is larger than a threshold value th1 (226), it is judged that a pixel of a coordinate (x, y) is sufficiently monotonously increased or decreased, and the value of the eval is increased by 1 (228). It is natural that natural dynamic images have an irregular variation, and the speed of the dissolve also becomes inconstant by the occurrence of unevenness in the case where a person performs the dissolve operation. Thus, a margin is given by providing the threshold value for judgement of monotony. The procedure returns to 208 and is repeated so that the above processes are performed for all pixels in the frame images (230 to 236). By this, the number of pixels satisfying the feature of the dissolve is put in the variable eval. Finally, it is checked whether or not the variable eval exceeds a threshold th2 (238), and if it exceeds, it is judged that the dissolve exists, and a dissolve detection process (240) is executed. Finally, the procedure returns to the process 204, and the processes from 204

are repeated to the end of the video image.

[0016] In the above method, even in the case where there is a movement of a camera, such as zoom or pan, the variable eval appears rather high. This is because if the camera moves, the brightness of each pixel in the frame image is also changed in response to that, and in such change, there are many pixels in which the brightness is monotonously increased or decreased. Thus, there is also a case where it is hard to differentiate between the dissolve and the movement of the camera. Then, in the following, such a dissolve detection method that the dissolve can be more clearly distinguished will be described.

[0017] In general, a time of the dissolve usually becomes 1 second (30 frames in the case of the NTSC system) or more. Accordingly, in a period when the dissolve is made, in a time of 22 frames when m = 8, or 15 frames even when m = 15, a state where the value of the variable eval is high continues. On the other hand, in the case of the movement of the camera, the value does not become high as in the dissolve, and a high state does not necessarily continuously continue. Accordingly, when the total sum of the values of the variable eval for the past n frames is taken, there appears a remarkable difference between the value of the sum in the dissolve and the sum in the movement of the camera. FIG. 4 shows a dissolve detection method in which the above approach is added.

[0018] First, as an initialization process, various variables necessary for the execution of a program are set to initial values

(400). Next, 0 is substituted in respective elements of m two-dimensional arrays B (x, y) containing brightness values of respective pixels of past frame images, and all n variables E1 to En storing values of the variable eval of past n frames are made 0 (402). When the size of the frame image is w × h, x takes a value from 0 to w-1, and y takes a value from 0 to h-1. In a process 404, the frame image outputted from the dynamic image reproduction device 10 is taken in (404). Hereinafter, processes 206 to 236 shown in FIG. 2 are executed to obtain the variable eval (406). Then, the value of the variable eval is substituted in En. The total of E1 to En is obtained in sum, and a shift is made while the value of Ej+1 is substituted in Ej in sequence, so that the newest eval value is always stored in E1 to En (408 to 412). Finally, it is judged whether the sum is larger than a threshold th3 (414), and if larger, the dissolve detection process 240 is performed, and if not, the procedure returns to the process 404 without performing anything, and is repeated.

[0019] In the dissolve detection process 240, a scene interposed between dissolves is selected as an important scene. When the dissolve detection methods of FIGS. 2 and 4 are executed, it is possible to obtain a graph expressing a time transition of an evaluation value like FIG. 5. The evaluation value does not instantaneously show a large value in the dissolve period, but has a feature showing a triangular change in which it is rapidly increased and is rapidly decreased. Then, two apexes constituting a bottom side of a triangle substantially

correspond to a start point and an end point of the dissolve. When a digest is prepared, if a portion where a special video effect is made, such as the dissolve, remains at the head or end, it is unsightly. Thus, a period 507 from a point where the dissolve is ended to a point before a next dissolve is started is cut out. For that purpose, in addition to a first threshold value 500 used for judgement of the dissolve in the above dissolve detection method, a second threshold value 502 lower than that is used. In the case where the dissolve as the start point of an important scene is detected, a point 506 when the evaluation value becomes first lower than the second threshold value after a point 504 when it exceeds the first threshold value, is made the start point of the important scene. At this time, the start point may be delayed for a margin. In the case where the dissolve is detected as the end point of the important scene, when the evaluation value is seen in the past from a point 510 when it exceeds the first threshold value, a point 508 when the evaluation value becomes first lower than the second threshold value is made the end point of the important scene. At this time, similarly to the start point, the end point may be made an early time for a margin. As the judgement whether the detected dissolve indicates the start point of the important scene or the end point, the time between dissolves can be used. If normal broadcasting continues, since there is no dissolve, a time interval between dissolves becomes long, and in the important scene, the interval is relatively short. By reproducing the thus obtained

important scenes in sequence, a digest is obtained.

[0020] In the above embodiment, although the monotonous change of brightness is checked, it is also possible to use a similar change of color. Differently from the brightness as one dimensional information, the color is three-dimension information. Accordingly, it is impossible to check a monotonous change simply on the basis of increase or decrease of values. Here, a simple change from color A to color B can be grasped as a tendency in which a distance from the color A is gradually increased, and a distance from the color B is gradually shortened when the two colors are mapped in a three-dimensional color space. Accordingly, instead of the two-dimensional array B storing the brightness values of the past frames in FIG. 2, a two-dimensional array B' storing colors is used, and if it is judged that the respective colors in B' are arranged in the form that a color difference from B'1 is increased and at the same time, a color difference from B'm is decreased, a method similar to the case of the brightness can be used thereafter.

(19)日本国特許庁(JP)　(12)公開特許公報（A）　(11)特許出願公開番号

特開平９－６５２８７

(43)公開日　平成9年(1997)3月7日

(51)Int. Cl.[6]	識別記号	庁内整理番号	F I		技術表示箇所
H04N	5/93		H04N	5/93	Z
G06T	13/00			5/268	
H04N	5/268		G06F	15/62 340	A

審査請求　未請求　請求項の数９　ＯＬ　（全１１頁）

(21)出願番号　特願平7-210409

(22)出願日　平成7年(1995)8月18日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地
(72)発明者　長坂　晃朗
東京都国分寺市東恋ケ窪1丁目280番地　株式会社日立製作所中央研究所内
(72)発明者　宮武　孝文
東京都国分寺市東恋ケ窪1丁目280番地　株式会社日立製作所中央研究所内
(72)発明者　藤田　武洋
東京都国分寺市東恋ケ窪1丁目280番地　株式会社日立製作所中央研究所内
(74)代理人　弁理士　小川　勝男

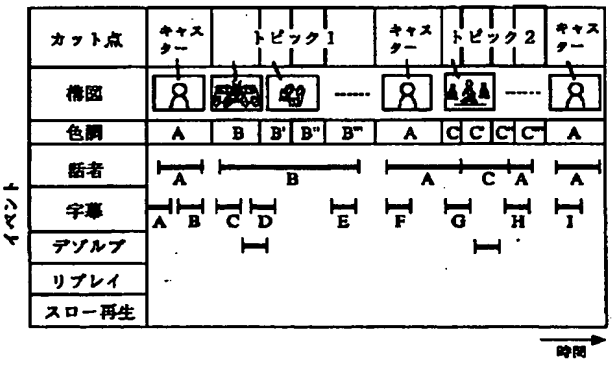
最終頁に続く

(54)【発明の名称】動画像の特徴場面検出方法及び装置

(57)【要約】
【目的】映像中の重要な場面かどうかの判定とその範囲の特定とを簡便かつ高速に行う。また，映像がどんな分野（ニュース，スポーツ中継等）に属するかを判定して分類し，ユーザの映像選択の一助となる情報として提供する。
【構成】対象となる動画像をフレーム単位で時系列に処理装置に入力する手段と，該処理装置において過去に入力されたフレームを複数枚バッファリングする手段と，該バッファリングされたフレームの特徴量が，古いものから順番に単調に最も新しいフレームの特徴量に近づいていく特徴を持つか否かを判定する手段とを設ける。さらに，該判定手段によって真と判定された場合には，特殊映像効果があったとして，そのフレームから一定時間もしくは次の特殊映像効果が検出されるまでの映像区間を重要な場面として抽出する手段を設ける。また，上記特殊映像効果に加えて各種の映像の変化や状態（カットの変わり目，字幕の表示，構図等）を検出する手段も設け，それらが現れる順序や組み合わせから映像の種類を判別する手段とを設ける。

図6

ニュース番組の典型的なイベントチャート



1

【特許請求の範囲】

【請求項１】連続する複数枚の画像よりなる動画像からデゾルブを含む特殊映像効果の場面の変わり目を検出する動画像の特徴場面検出方法において，対象となる動画像をフレーム単位で時系列に処理装置に入力し，該処理装置では，フレーム中の各画素の色もしくは輝度が，連続する複数枚のフレーム群にまたがって，該フレーム群の最初のフレームの色もしくは輝度の値から，最後のフレームの色もしくは輝度の値に向けて単調に近づく傾向で推移しているかどうかを調べ，該条件を満たす画素の数から画面全体としての変化を表す評価値を計算し，該評価値が予め定めた許容範囲外となった時点で，該連続する複数枚のフレーム群にまたがる区間に，デゾルブを含む特殊映像効果による場面の変わり目があったと判定し，該区間もしくはその近傍を動画像中の特徴的な点であると判定することを特徴とする動画像の特徴場面検出方法。

【請求項２】請求項１記載の動画像の動画像の特徴場面検出方法において，単調に近づく傾向かどうかの判定の手段として，各画素ごとに複数枚のフレーム分の輝度を格納するバッファを持たせ，該バッファ内において直前の輝度との差分が正の値を示す輝度の数が該バッファに格納された輝度の総数のうち多数を占める場合に単調増加とし，一方，該バッファ内において直前の輝度との差分が負の値を示す輝度の数が該バッファに格納された輝度の総数のうち多数を占める場合に単調減少と判定することを特徴とする動画像の特徴場面検出方法。

【請求項３】請求項１記載の動画像の動画像の特徴場面検出方法において，画面全体としての時間推移の傾向を表す評価値を一定時間分過去まで遡って総和をとり，その総和をもってデゾルブを含む特殊映像効果による場面の変わり目があったと判定することを特徴とする動画像の特徴場面検出方法。

【請求項４】請求項１記載の動画像の動画像の特徴場面検出方法において，検出したデゾルブを含む特殊映像効果による場面の変わり目と後続するデゾルブを含む特殊映像効果による場面の変わり目に挟まれた区間が一定時間以内であるとき、該区間を重要な場面として抽出することを特徴とする動画像の特徴場面検出方法。

【請求項５】請求項４記載の動画像の動画像の特徴場面検出方法において，該重要な場面の区間は、特殊映像効果の継続期間中を除いた残りの区間とすることを特徴とする動画像の特徴場面検出方法。

【請求項６】連続する複数枚の画像よりなる動画像がどのような種類の番組であるかを判別する動画像の分類方法において，対象となる動画像をフレーム単位で時系列に処理装置に入力し，または対象となる音声を時系列に処理装置に入力し，該処理装置では，カット変化や色調を含む複数の種類の画像特徴量の変化を検出する手段と，必要に応じて話者変化を含む音声特徴量の変化を検出する手段を設け，該画像特徴量および音声特徴量の変化検出手段により，変化が発生したこと，もしくは複数の変化が同時または特定の順番で発生したことからなる特徴量に基づき，番組の種類を判別することを特徴とする動画像の分類方法。

2

【請求項７】請求項６記載の動画像の分類方法において，検出された複数の種類の画像特徴量もしくは音声特徴量の変化点ならびにその変化区間を，時間軸を１つの軸とする表形式で一覧表示することを特徴とする動画像の分類方法。

【請求項８】連続する複数枚の画像よりなる動画像からデゾルブを含む特殊映像効果の場面の変わり目を検出する動画像の特徴場面検出装置において，対象となる動画像をフレーム単位で時系列に入力する入力手段と、過去に入力されたフレームを複数枚バッファリングする手段と，該バッファリングされたフレームの特徴量が，古いものから順番に単調に最も新しいフレームの特徴量に近づいていく特徴を持つか否かを判定する手段と、該判定手段によって真と判定された場合には，特殊映像効果があったとして，そのフレームから一定時間もしくは次の特殊映像効果が検出されるまでの映像区間を重要な場面として抽出する手段を設けたことを特徴とする動画像の特徴場面検出装置。

【請求項９】請求項９記載の動画像の特徴場面検出装置において、上記特殊映像効果に加えて、カットの変わり目，字幕の表示，構図を含む映像の変化や状態を検出する手段と、それらが現れる順序や組み合わせから映像の種類を判別する手段とを設けたことを特徴とする動画像の特徴場面検出装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は，ビデオや映画等の動画像を短時間で概要把握を行うための早見する方法及び装置に係り，特にビデオテープやビデオディスクに格納された動画像からカット（１台のカメラで撮影された途切れのない動画像区間）間のデゾルブ（連続するカットA，Bがあるとき，そのカットの変わり目において，Aがフェードアウトすると同時にBがフェードインする特殊映像効果）を検出することによって動画像を代表する場面を特定する動画像の特徴場面検出方法及び装置に関する。

【０００２】

【従来の技術】近年，通常のテレビ放送に加えて，衛星放送やケーブルテレビなどが普及しつつあり，放送の多チャンネル化が進行している。今後，情報ハイウエイと称される広帯域の通信基盤が整備されれば，放送の配信が容易になり，現状よりもさらに多くの放送業者が参入して，多チャンネル化が加速されると考えられる。こうした大量に放送される情報の中から，視聴者個人個人にとって有用な情報と無用な情報とを区別し，選択するこ

とは非常に手間と時間のかかる作業である。そのため，映像内容を手早く把握するための要約情報（ダイジェスト）を効率よく作成する技術の研究が進められている。ダイジェストを作成するにあたって最も基本的かつ不可欠な処理は，映像中から重要な場面を選び出すことである。もし，映像中の場面場面の重要度を計算機で自動的に判定できれば，ダイジェストの作成は非常に簡単になる。例えば，特開平4-294694号では，野球中継において，映像中の移動物体の移動結果と，ある特定のイベントとの対応（ランナーの本塁位置への移動と，得点があったこととの対応等）に着目して，重要度の高い場面を選択する方法が示されている。

【０００３】

【発明が解決しようとする課題】しかしながら，移動物体の動き解析は，現状の画像認識の技術水準では精度や処理速度が十分でなく，それによって得られた動きパターンと，特定のイベントとの対応が必ずしも対応するとは限らないという問題点がある。また，正しくイベントが検出できた場合でも，その前後のどの範囲までを重要な場面として切り出せばよいのかを自動判定させることは極めて困難である。さらに，ダイジェスト自体，映像全体を視聴するのに比べれば格段に短い時間ながら，やはり一定の時間をかけて視聴する必要性は残っており，もっと簡潔に概要把握できるような技術が求められている。

【０００４】本発明の目的は，映像中の重要な場面かどうかの判定とその範囲の特定とを簡便かつ高速に行うための方法を提供することにある。また，映像がどんな分野（ニュース，スポーツ中継等）に属するかを判定して分類し，ユーザの映像選択の一助となる情報として提供することにある。

【０００５】

【課題を解決するための手段】放送映像については，多くの場合，放送局側で重要な場面を強調するような各種の映像効果が施されている。この性質はスポーツ中継の場合に特に顕著であり，例えば，得点が入った場合にはリプレイを放映するといった特徴がある。リプレイ映像は視点の異なるカメラで撮像された映像が使われることが多く，単純に全く同じ映像かどうかでリプレイ映像か否かを判定することはできないが，そうしたリプレイ映像に切り替わるときには，デゾルブやワイプといった特殊映像効果が用いられ，通常の放送から一時的に外れることを視聴者が明確に分かるような工夫がされている。さらにまた通常の放送に戻るときにも同様の映像効果が利用される。したがって，こうした特殊映像効果を検出することにより，重要な場面を選び出すことが可能になる。

【０００６】そこで，対象となる動画像をフレーム単位で時系列に処理装置に入力し，該処理装置では，フレーム中の各画素の色もしくは輝度が，連続する複数枚のフレーム群にまたがって，該フレーム群の最初のフレームの色もしくは輝度の値から，最後のフレームの色もしくは輝度の値に向けて単調に近づく傾向で推移しているかどうかを調べ，該条件を満たす画素の数から画面全体としての変化を表す評価値を計算し，該評価値が予め定めた許容範囲外となった時点で，該連続する複数枚のフレームにまたがる区間に，デゾルブ等の特殊映像効果による場面の変わり目があったと判定し，該区間もしくはその近傍を動画像中の特徴的な点であると判定する。

【０００７】また，対象となる動画像をフレーム単位で時系列に処理装置に入力し，または対象となる音声を時系列に処理装置に入力し，該処理装置では，カット変化や色調を含む複数の種類の画像特徴量の変化を検出する手段と，必要に応じて話者変化を含む音声特徴量の変化を検出する手段を設け，該検出手段により，変化が発生したこと，もしくは複数の変化が同時または特定の順番で発生したことからなる特徴量に基づき，番組の種類を判別する。

【０００８】

【作用】放送でリプレイされる場面は，専門家が重要であると判定した部分であり，そうしたリプレイ場面を検出できれば，ダイジェスト作成が極めて容易になる。本発明によれば，デゾルブを含む特殊映像効果による場面の変わり目が検出できるため，そうした特殊効果に相前後して流される重要な場面を精度よく抽出できる。また同時に，その場面の範囲も得ることができる。

【０００９】さらに，カット変化や色調を含む複数の種類の画像特徴量の変化が同時または特定の順番で発生したことからなる特徴量に基づき，番組の種類を判別する手段によって，映像の種類が自動的に判定されるので，視聴者にとって興味のない種類の映像であれば，ダイジェスト映像を見るまでもなく却下でき，効率的な映像選択ができる。また，この映像の種類の判定においては，簡単な画像や音声の変化とその組み合わせから判定を行うので，処理が高速に行える。

【００１０】

【実施例】以下，本発明の１実施例を詳細に説明する。

【００１１】図１は，本発明を実現するためのシステム構成の概略ブロック図の一例である。１はＣＲＴ等のディスプレイ装置であり，コンピュータ４の出力画面を表示する。コンピュータ４に対する命令は，キーボードやポインティングデバイス等の入力装置５を使って行うことができる。１０の動画像再生装置は，地上波放送や衛星放送，ケーブルテレビなどの放送番組を受信するためのチューナー装置，もしくは光ディスクやビデオテープ等に記録された動画像を再生するための装置である。動画像再生装置から出力される映像信号は，逐次，３のＡ／Ｄ変換器によってデジタル画像データに変換され，コンピュータに送られる。コンピュータ内部では，デジタル画像データは，インタフェース８を介してメモリ９に入

り，メモリ９に格納されたプログラムに従って，ＣＰＵ７によって処理される。１０が扱う動画像の各フレームに，動画像の先頭から順に番号（フレーム番号）が割り付けられている場合には，フレーム番号を制御線２によって動画像再生装置に送ることで，当該場面の動画像を呼び出して再生することができる。また，処理の必要に応じて，各種情報を６の外部情報記憶装置に蓄積することができる。メモリ９には，以下に説明する処理によって作成される各種のデータが格納され，必要に応じて参照される。

【００１２】以下では，重要場面の選別にあたって，特殊映像効果によるカット変化の一つであるデゾルブを検出する方法について詳細に説明する。

【００１３】図２は，図１で示したシステム上で実行される，動画像のデゾルブ検出プログラムのフローチャートの一例である。プログラムはメモリ９に格納され，ＣＰＵ７は，まず最初に初期化処理として，プログラムの実行に必要な各種の変数を初期値に設定する（２００）。次に，過去のフレーム画像の各画素の輝度値を収めるｍ個の二次元配列B(x, y)の各要素に０を代入する（２０２）。フレーム画像のサイズがｗ×ｈのとき，ｘは０からw-1，ｙは０からh-1までの値をとる。処理２０４では，動画像再生装置１０が出力するフレーム画像の取り込みを行う（２０４）。処理２０６は，評価値が入る変数evalを０にし，ループカウンタに初期値０を代入する。そして，以下の２０８～２２８の処理をフレーム画像中の全画素について行う。

【００１４】２０８から２２８の処理では，デゾルブに特有の性質の検出を行っている。ここで，デゾルブは，図３に示すように，カットの変わり目の前後でＢのように，前後のカットのフレーム画像ＡとＣとが混じりあう区間を持つカット変化である。ＢにおけるＡとＣの混合比率は，デゾルブ開始時のＡが１００％，Ｃが０％の状態から，時間をかけて比率が逆転してゆき，最終的にＡが０％，Ｃが１００％になった時点でデゾルブが完了する。濃淡画像の場合，Ａの輝度値をBa，Ｂの輝度値をBb，Ｃの輝度値をBc，Ｃの混合割合をα（$0 \leqq \alpha \leqq 1$）としたとき，$Bb = Ba \times (1 - \alpha) + Bc \times \alpha$の式で近似することができる。この式を変形すると，$Bb = (Bc - Ba) \times \alpha + Ba$になり，混合割合$\alpha$が０から単調に増加するデゾルブの場合，Bbの値もBaからBcまで単調に増加もしくは減少する。したがって，過去ｍフレーム分について常に画素の輝度値をバッファに蓄えておき，そのｍフレーム長の区間で輝度値が単調に増加もしくは減少しているかどうかを調べることでデゾルブの検出を行うことができる。ｍの値は，８から１５程度に設定すると，実験的に良好な結果が得られる。

【００１５】まず処理２０８では，過去のフレームの輝度値を記憶している二次元配列Ｂのｍ番目の配列Bmに，座標(x, y)で表される画素の輝度値を代入する。そして，ループカウンタiに１を代入し，変数numに０を代入する。次に，１番目の配列に記憶された輝度値Bl(x, y)とｍ番目の配列Bm(x, y)の値を比較し（２１２），続けて，i番目の配列に記憶された輝度値Bi(x, y)がその次の配列Bi+1(x, y)の値よりも大きいかどうかを比較する（２１４，２１６）。Bl(x, y)がBm(x, y)より大きいときには，Bi(x, y)がBi+1(x, y)より大きい場合にnumの値を１つ増やす。逆に，Bl(x, y)がBm(x, y)より小さいときには，Bi(x, y)がBi+1(x, y)より小さい場合にnumの値を１つ増やす（２１８）。続く処理２２０では，Bi(x, y)にBi+1(x, y)の値を代入することで，ｍ個の配列Ｂを順番に１つずつシフトするようにし，常に最新のフレームから数えてｍフレーム分の輝度値がバッファとして格納されているようにする。処理２２２では，ループカウンタiを１つ増やし，iがｍより大きくなるまで，処理２１２の時点でBl(x, y)がBm(x, y)より大きかったときには処理２１４，そうでないときには処理２１６に戻って処理を繰り返す（２２４）。numが閾値th1よりも大きいときには（２２６），座標(x, y)の画素については，十分単調に増加もしくは減少しているとしてevalの値を１つ増やす（２２８）。自然動画像はノイズ等により不規則な変動があるのが常であり，また，デゾルブの速度も，人間がデゾルブ操作を行う場合にはムラが生じて一定ではなくなるので，単調性の判定に閾値を設けることでマージンを持たせる。上記処理をフレーム画像中の全画素について行うべく，２０８に戻って繰り返す（２３０～２３６）。これによって，デゾルブの特徴を満たす画素の数がevalに入る。最後に，evalが閾値th2を超えているかどうかを調べ（２３８），超えていればデゾルブがあるとして，デゾルブ検出処理（２４０）を実行する。最後に，処理２０４に戻り，映像の終わりまで２０４からの処理を繰り返す。

【００１６】上記の方法では，ズームやパンといったカメラの動きがある場合にも，evalが高めに出る。カメラが動けば，それに応じて，フレーム画像中の各画素の輝度も変化し，そうした変化の中には，輝度が単調増加もしくは単調減少している画素も少なからず存在するからである。そのため，デゾルブとカメラの動きとの区別がつきにくいケースもある。そこで，以下では，デゾルブがもっと明確にわかるようなデゾルブ検出方法について説明する。

【００１７】一般に，デゾルブの時間は，１秒（NTSC方式の映像の場合で３０フレーム）以上になるものが多い。したがって，デゾルブがかかっている区間では，ｍ＝８のときで２２フレーム，ｍ＝１５のときでも１５フレーム以上の時間，evalの値が高い状態が続く。一方，カメラの動きの場合は，デゾルブのときほど値は高くない上，必ずしも連続して高い状態が続くとは限らない。したがって，過去ｎフレーム分についてevalの値の総和sumをとったとき，デゾルブのときのsumの値とカメラの

動きのときのsumとでは顕著な違いが現れる。図４は，上記の考え方を加えたデゾルブ検出方法である。

【００１８】まず最初に初期化処理として，プログラムの実行に必要な各種の変数を初期値に設定する（４００）。次に，過去のフレーム画像の各画素の輝度値を収めるｍ個の二次元配列B(x, y)の各要素に０を代入するとともに，過去ｎフレーム分のevalの値を記憶するｎ個の変数E1～Enを全て０にする（４０２）。フレーム画像のサイズがｗ×ｈのとき，ｘは０からw-1，ｙは０からh-1までの値をとる。処理４０４では，動画像再生装置１０が出力するフレーム画像の取り込みを行う（４０４）。以下，図２で示した２０６から２３６までの処理を実行してevalを得る（４０６）。そして，Enにevalの値を代入する。E1からEnまでの総和をsumに求めるとともに，EjにEj+1の値を次々と代入しながらシフトし，常に最新のeval値がE1～Enに格納されているようにする（４０８～４１２）。最後に，sumが閾値th3よりも大きいかどうかを判定し（４１４），大きければ，デゾルブ検出処理２４０を行い，そうでなければ何もせずに処理４０４まで戻って繰り返す。

【００１９】デゾルブ検出処理２４０では，デゾルブで挟まれた場面を重要な場面として選択する。図２および図４のデゾルブ検出方法を実行すると，図５のような評価値の時間推移を表すグラフを得ることができる。評価値は，デゾルブ区間において，一瞬だけ大きな値を示すのではなく，急速に増加して急速に減少する三角形状の変化を示す特徴がある。そして，三角形の底辺を成す２頂点が，デゾルブの開始点と終了点にほぼ対応している。ダイジェストを作成するときには，デゾルブのような特殊映像効果がかかった部分が先頭や末尾に残っていると見苦しいので，デゾルブの終わった点から，次のデゾルブが始まる手前までの区間５０７を切り出すようにする。そのため，上記のデゾルブ検出方法でデゾルブか否かの判定に用いる第１の閾値５００に加えて，それより低い第２の閾値５０２を用いる。そして，重要場面の開始点としてのデゾルブが検出された場合には，評価値が第１の閾値を超えた点５０４以降ではじめて第２の閾値を下回った点５０６を重要場面の開始点とする。このとき，余裕をとって開始点を遅らせても構わない。また，重要場面の終了点としてデゾルブが検出された場合には，評価値が第１の閾値を超えた点５１０から過去に遡って見たときに初めて第２の閾値を下回った点５０８を重要場面の終了点とする。このとき，開始点と同様に，余裕をとって終了点を早めの時間にとってもよい。検出されたデゾルブが重要場面の開始点を示すのか，終了点を示すのかの判定には，デゾルブ間の時間が利用できる。通常の放送が続いてれば，デゾルブはないのでデゾルブ間の時間間隔が長くなり，重要場面ならば，比較的間隔は短い。こうして得られた重要場面を順番に再生することで，ダイジェストができる。

【００２０】上記の実施例においては，輝度の単調な変化を調べたが，色の同様の変化を利用することもできる。色は１次元情報である輝度と異なり，３次元の情報である。従って，単純に値の増加減少をもとに単調変化を調べることはできない。ここで，A色からB色への単調な変化とは，２つの色を３次元の色空間にマッピングしたとき，A色からの距離を徐々に増しつつ，B色との距離を徐々に縮める傾向としてとらえることができる。したがって，図２における過去のフレームの輝度値を記憶する二次元配列Ｂの替わりに，色を記憶する二次元配列Ｂ’を用い，そのＢ’中の各色がＢ’１との色差が増加すると同時にＢ’ｍとの色差が減少する形で並んでいることを判定すれば，あとは輝度の場合と同様の手法を用いることができる。

【００２１】上記のようなデゾルブ等の特殊映像効果を使ったシーンを重要場面とみなせるのは，現実としてスポーツ中継等の一部の番組に限定される。また，スポーツ番組中でも合間に挿入されるコマーシャル中には特殊映像効果が頻繁に登場するため，単純にデゾルブに挟まれた区間という条件では過剰に検出しすぎることも多い。もちろん，多めに検出する分には，元の映像よりも十分に短い映像になっていれば，実用上問題はない。しかし，より精度高く重要場面を抽出できれば，概要把握にかかる時間がさらに節約できる。そこで，ダイジェストを作成する対象の映像がどのような種類の映像かを区別する手段を設け，重要場面の選択に活用する。

【００２２】図６と図７は，それぞれニュース番組とスポーツ番組において発生するイベントを時間軸に沿って図示したものである。ここでは，イベントとして，画像や音声の特徴が大きく変化する点を考える。図中では，１）構図，２）色調，３）話者，４）字幕，５）デゾルブ，６）リプレイ，７）スロー再生，の７項目を例に挙げた。こうしたイベントの現れ方や組み合わせには番組の種類によって特徴があり，その特徴をもとに番組の分類を行うことができる。例えば，ニュース番組においては，キャスターが全面に登場するカットが時間を空けて複数回現れるので，同じ構図の画像，より具体的には中心付近に顔の色である肌色が大きな面積を占めている画像が複数回現れる特徴がある。また，そのときの話者は同一人物である場合が多いとか，番組全体として字幕が頻繁に現れるという特徴もある。一方，スポーツ中継の場合，固定位置に設置された複数のカメラを切り替えながら放送が行われることが多く，同じか極めて類似した構図の画像が頻繁に現れる。特に野球やサッカーの場合には，色調は芝生の色である緑がメインとなる。また，リプレイやスロー再生が頻繁に使われるという特徴がある。さらに，ＣＭの場合には，音の途切れが少ない，ＢＧＭが頻繁に使われる，色調が鮮やか，カットが多く，その時間長も短い，などの特徴がある。このように，映像中における複数のイベントの組み合わせパターンか

ら，その映像の種類をある程度推測することができる。そして，ここで挙げたイベントは，画像認識・音声認識の技術を要する中では比較的簡単に求められ，その信頼性が高いものばかりである。すなわち，ストーリー等の映像の意味内容に関する認識は必要としない。

【００２３】図８は，映像の種類を見分けるシステムのブロック図の一例である。入力映像は，画像信号と音声信号のそれぞれについて，画像取り込み部８００及び音声取り込み部８０２でデジタイズされる。デジタイズされたデータは，イベント検出部８０４に送られ，８０４中の種類別に設けられた専用検出部８０６～８２０によって，イベント検出の処理が行われる。検出されたイベントは，イベント別カウンタ部８２２によって，イベントの種類別にカウントされる。また，同時生起カウンタ部８２４は，複数のイベントが同時に，もしくは規定の順番に現れた場合にのみ，そのイベントの組み合わせに対応するカウンタを１増やす。これらのカウンタで得られた各種イベントの出現頻度分布は，比較部８２８によって，どの種類の番組におけるイベントの出現頻度分布に近いか比較照合される。

【００２４】次に，図８中の各ブロックについて詳細に説明する。

【００２５】イベント検出部８０４のうち，カット点検出部８０６は，カットの変わり目を検出する。その手法については，例えば，発明者らによる，情報処理学会論文誌Vol.33，No.4，「カラービデオ映像における自動索引付け法と物体探索法」や特開平４－１１１１８１号等で示された方法等が利用できる。イベント別カウンタ部８２２では，カット点の数がカウントされる。

【００２６】同一構図検出部８０６は，予め定めた時間以内の過去に遡って，同じ構図もしくは類似した構図の絵が現れているかどうかを検出する。これにはテンプレートマッチングに代表される画像比較手法が使える。具体的には，比較する２枚のフレーム画像の同じ座標位置にある画素の１つ１つについて，輝度差もしくは色差を求めて全画面分の総和をとり，これを画像間の相異度とする。この相異度が定めた閾値より小さければ，同一もしくは類似性が高いと判定できる。ここで，映像中のフレーム画像全てについて，同一構図か否かを検出するのは処理時間がかかり，また，連続するフレーム画像間では画像の類似性が高い動画像の特徴を考慮すると無駄でもある。そこで，カット点検出に連動させて，カット点の画像だけを調べる対象とする。イベント別カウンタ部では，同一構図を持つフレームの数がカウントされる。

【００２７】色調検出部８１０は，予め定めた時間以内の過去に遡って，同一の色調もしくは類似した色調の絵が現れているかどうかを検出する。これには，例えば，フレーム画面全体についての色度数分布が利用できる。これは構図に無関係な，どの色がどれだけ使われているかを表した特徴量である。具体的には，比較する２枚のフレーム画像のそれぞれについて，画像を表現する画素の色を６４色程度に分別し，それら各色がそれぞれフレーム画像中にどれだけ存在するかをカウントする。そして，得られた度数分布の各度数の差分の絶対値の総和をもって色調の相異度とする。この相異度が定めた閾値より小さければ，同一もしくは類似性が高いと判定できる。色調に関しても構図と同様の理由で，カット点の画像についてのみ対象とすると効率がよい。イベント別カウンタ部では，同一色調を持つフレームの数がカウントされる。また，色調検出部は，途中で求めた度数分布を利用して，どの色が最も多く使われているかを調べるようにしてもよい。具体的には，イベント別カウンタ部中に，赤・青・緑等の色別にカウンタを用意し，赤系の色が多ければ赤のカウンタを増やし，緑が多ければ，緑のカウンタを増やすようにする。

【００２８】字幕検出部８１２は，映像中に字幕が現れているかどうかを検出する。その手法については，例えば，発明者らによる，特願平5-330507等で示された方法等が利用できる。イベント別カウンタ部８２２では，字幕の出現数がカウントされる。

【００２９】デゾルブ検出部８１４は，映像中のデゾルブ等の特殊効果を検出する。その手法については，本発明の前半で説明した通りである。イベント別カウンタ部８２２では，デゾルブの出現数がカウントされる。

【００３０】リプレイ検出部８１６は，予め定めた時間以内の過去に遡って，全く同一の映像が現れているかどうかを検出する。これは同一構図検出部８０８と同様にテンプレートマッチング等によってフレーム画像の比較をすることで行える。しかし，比較する動画像間の各フレームごとにテンプレートマッチングを行っていたのでは処理時間がかかりすぎるので，各フレームを数文字分程度のコードに変換し，そのコード列の照合をもって動画像の照合とする。１枚のフレームに対応するコード単体では情報量が極めて小さいが，動画像は多くのフレームから構成されるので，１つの動画像が含むコードの数も多く，動画像中におけるコードの一連のシーケンスは，一片の動画像を特定するに足る十分な情報量を持つ。こうした考え方に立脚した動画像の照合方法は，発明者らによる，特開平７－１１４５６７号に示されている。

【００３１】スロー再生検出部８１８は，スロー再生の映像を検出する。スロー再生は，フレーム画像を標準再生時よりも長めの間隔（１／２スローで２倍，１／４スローで４倍）で連続表示することで実現されるため，スロー再生の映像の場合，画像取り込み部８００でデジタイズされる画像は，全く同じ画像が複数枚続くという特徴がある（１／２スローで２枚，１／４スローで４枚）。そこで，スロー再生かどうかの判定には，連続する２枚のフレームを調べ，そのテンプレートマッチングによって画像相異度を調べる。そして，一定時間分の相

異度の推移を調べ，相異度が特定の周期で大きい値と小さい値を繰り返しているようならば，スロー再生であると判定する。例えば，１／２スローの場合には，２枚ずつ同じ画像が続くので，相異度は，小さい値と大きい値を交互に繰り返す。１／４の場合には，小さい値が３回続いて大きい値が１回というように繰り返す。但し，動画像の場合，スロー再生でなくても，連続する２枚のフレーム画像は類似しているので，相異度の大小の判定は閾値を低めにして行う必要がある。イベント別カウンタ部８２２では，スロー再生の出現数がカウントされる。

【００３２】同一話者検出部８２０では，予め定めた時間以内の過去に遡って，同一の話者が話したことがあったかどうかを検出する。例えば，音声の自己相関を求め，最も大きな値をとる周波数が一致しているかどうかで調べることができる。イベント別カウンタ部８２２では，同一話者の発話数がカウントされる。

【００３３】同時生起カウンタ部８２４は，上記のイベントのうちの幾つかが同時もしくは特定の順番で現れた場合にカウントを行う。カウンタは，検出するイベントの組み合わせの数だけ用意される。例えば，同じ構図のときに，同じ話者が話しているケースでは，構図イベントと話者イベントの同時発生に対応するカウンタが１増やされる。同様に，デゾルブがあって，その直後にスロー再生が検出された場合には，デゾルブイベントとスロー再生イベントの連続発生に対応するカウンタが１増える。

【００３４】比較部８２８では，時計８２６を参照し，時刻ｔ１からｔ２までの一定時間における映像中のイベントの出現頻度の傾向が，どのような種類の番組のものに近いかを比較する。比較に先立ち，まずニュース番組，スポーツ番組などそれぞれの種類別に典型的なイベントを調べておき，番組を特徴づける重要なイベントであるほど高くなるように値を与えてランク付けを行って，番組ごとにイベント別のランク一覧表を作成する。比較にあたっては，各イベントの出現頻度値を正規化した値に，このランク一覧表で記述された値を掛けて重み付けを行い，そうして得られた各イベントごとの値の総和が閾値を超えた場合，そのランク一覧に対応する種類の番組であると判定する。

【００３５】このようにして得られたイベントを，図６もしくは図７のような，一方を時間軸とする表形式で，図１のディスプレイ１上に一覧表示することができる。この一覧表示によって，計算機が自動で判定できなかった場合でも，ユーザはこうした情報を１つの手がかりにして，他から入手した情報，経験や知識等を合わせて利用することによって，番組の種類を推測できる可能性がある。また，計算機に教えていない種類の番組が新たに入力された場合，この一覧表示の中から，重要なイベント，もしくはイベントの組み合わせを選んで登録するようにしてもよい。これは，図１で示したマウス等のポインティングデバイス５を使って，一覧表上の各イベントの変化点や区間の表示部分をクリックするなどのダイレクトかつビジュアルな操作で行うようにすればユーザにとって非常に便利になる。

【００３６】尚、本発明はＰＣ／ＷＳを用いて実現できる他、ＴＶ、ＶＴＲなどの一機能としても適用可能である。

【００３７】

【発明の効果】本発明によれば，重要な場面とその範囲を同時に得ることができ，ダイジェスト映像が自動で作成できる効果がある。一般にリプレイされる場面は重要な場面であることが多いが、本発明では、デゾルブを含む特殊映像効果の区間を検出することによって、放送中のリプレイ場面を精度よく検出できる。

【００３８】さらにまた，カット変化や色調を含む複数の種類の画像特徴量の変化が同時または特定の順番で発生したことからなる特徴量に基づき，番組の種類を判別する手段によって，映像の種類が自動的に判定されるので，視聴者にとって興味のない種類の映像であれば，ダイジェスト映像を見るまでもなく却下でき，効率的な映像選択ができる効果がある。また，この映像の種類の判定においては，簡単な画像や音声の変化とその組み合わせから判定を行うので，処理が高速に行える。

【図面の簡単な説明】

【図１】本発明の実施例を実現するためのシステムブロック図である。

【図２】デゾルブの検出を行うプログラムのフローチャートである。

【図３】デゾルブの概念を表す図である。

【図４】デゾルブの検出を行うもう１つのプログラムのフローチャートである。

【図５】デゾルブ検出を行うプログラムを実行したときの評価値の時間推移を表すグラフである。

【図６】ニュース番組の典型的なイベントチャートである。

【図７】スポーツ中継の典型的なイベントチャートである。

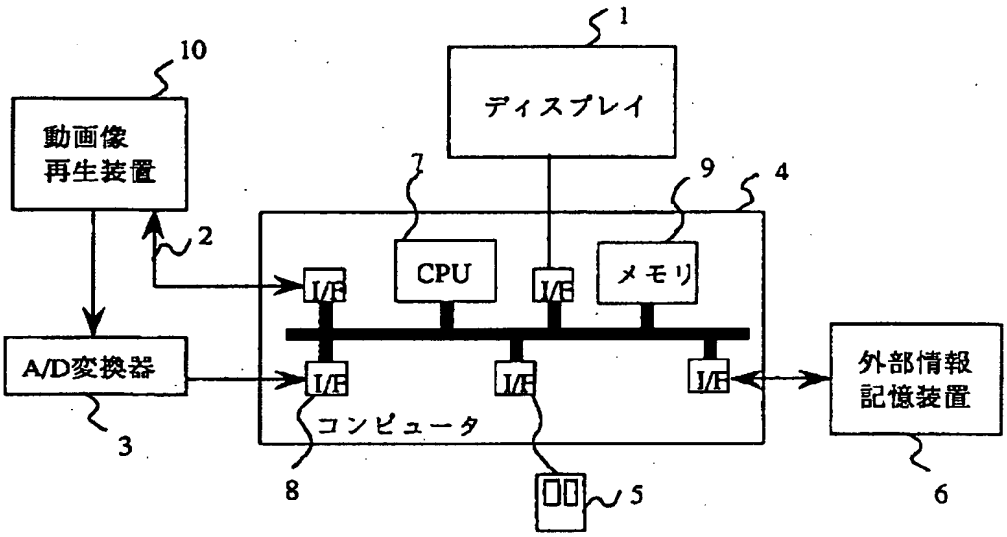
【図８】映像の分類を行うシステムのブロック図である。

【符号の説明】

１…ディスプレイ，２…制御信号線，３…Ａ／Ｄ変換器，４…コンピュータ，５…入力装置，６…外部情報記憶装置，７…ＣＰＵ，８…接続インタフェース，９…メモリ，１０…動画像再生装置，１１…キーボード。

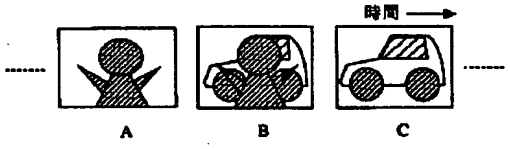
【図１】

図１



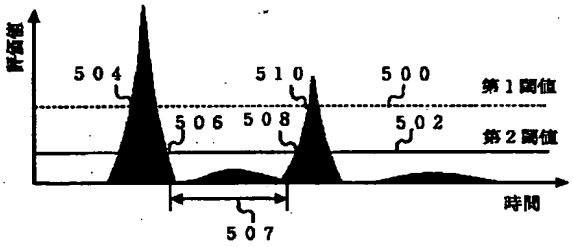
【図３】

図3



【図５】

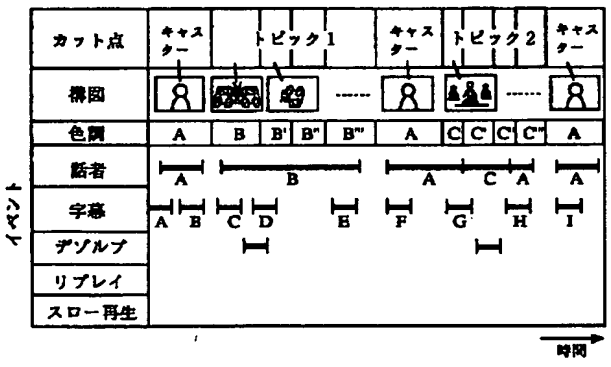
図5



【図６】

図6

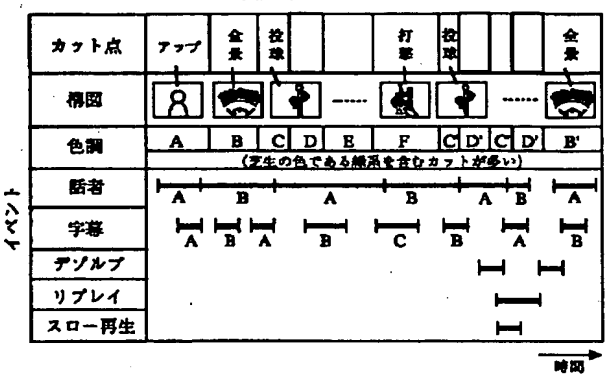
ニュース番組の典型的なイベントチャート



【図７】

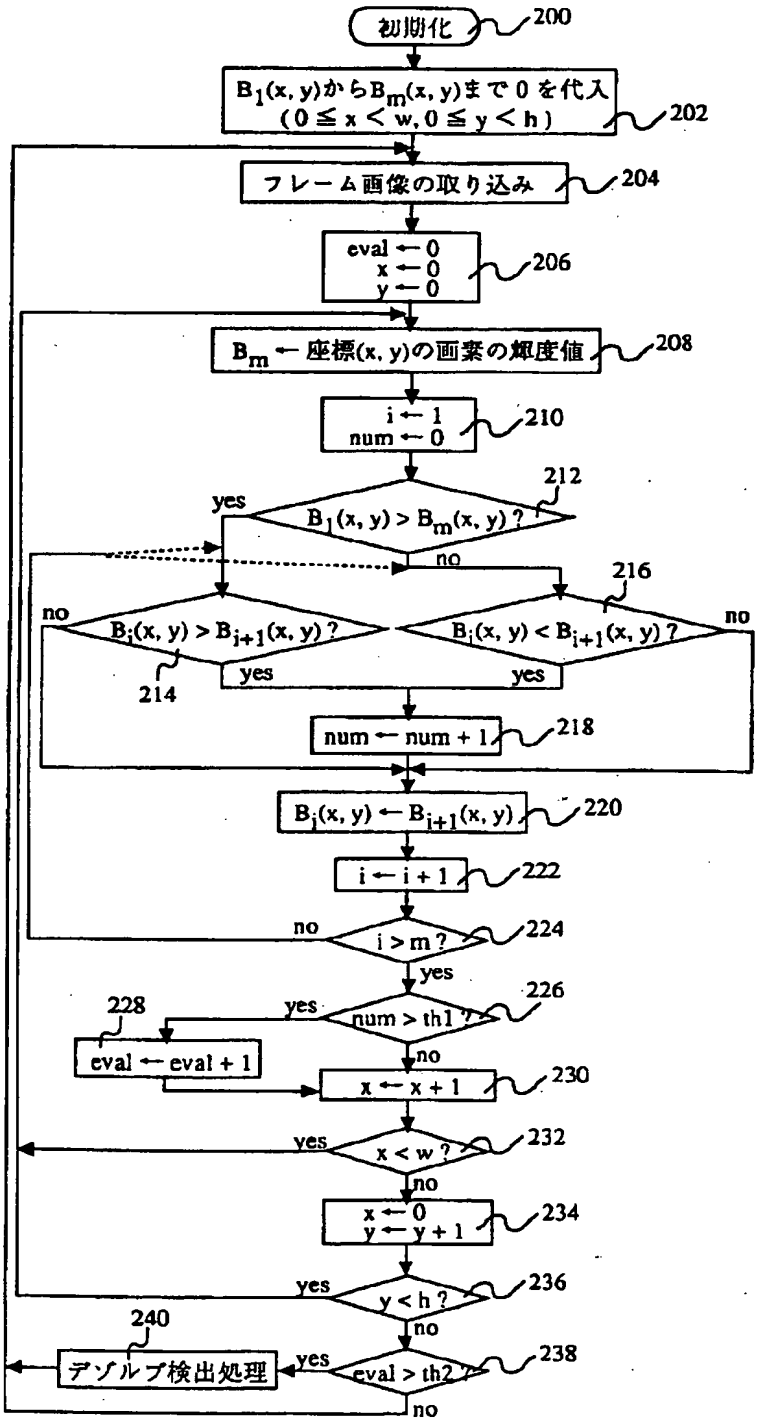
図7

スポーツ中継の典型的なイベントチャート



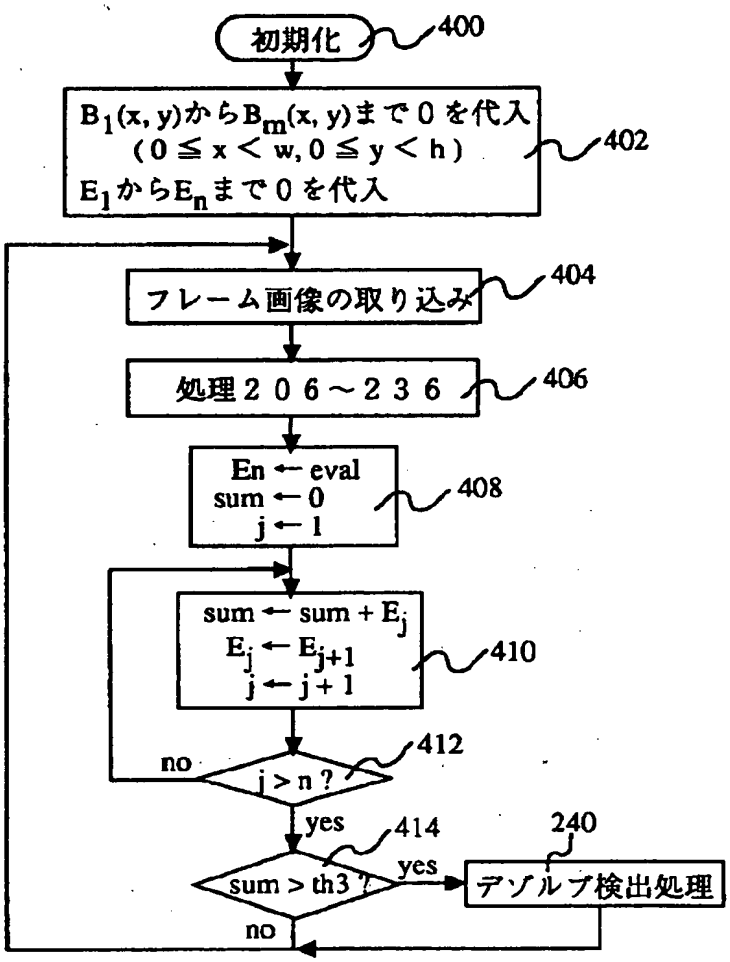
【図２】

図２


初期化
200
$B_1(x, y)$から$B_m(x, y)$まで０を代入
$(0 \leqq x < w, 0 \leqq y < h)$
202
フレーム画像の取り込み
204
eval ← 0
x ← 0
y ← 0
206
B_m ← 座標(x, y)の画素の輝度値
208
i ← 1
num ← 0
210
212
yes
$B_1(x, y) > B_m(x, y)$?
no
216
no
$B_i(x, y) > B_{i+1}(x, y)$?
$B_i(x, y) < B_{i+1}(x, y)$?
no
214
yes
yes
num ← num + 1
218
$B_i(x, y) \leftarrow B_{i+1}(x, y)$
220
i ← i + 1
222
no
i > m ?
224
yes
228
yes
num > th1 ?
226
eval ← eval + 1
no
x ← x + 1
230
yes
x < w ?
232
no
x ← 0
y ← y + 1
234
yes
y < h ?
236
no
240
デゾルブ検出処理
yes
eval > th2 ?
238
no

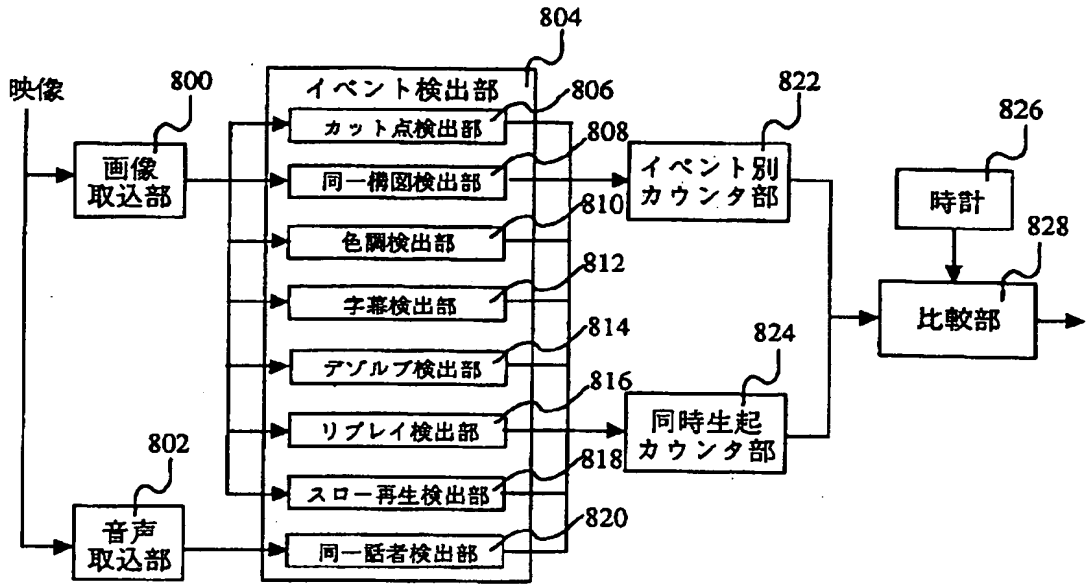
【図４】

図４


初期化
400
$B_1(x, y)$から$B_m(x, y)$まで０を代入
$(0 \leq x < w, 0 \leq y < h)$
E_1からE_nまで０を代入
402
フレーム画像の取り込み
404
処理２０６～２３６
406
En ← eval
sum ← 0
j ← 1
408
sum ← sum + E_j
E_j ← E_{j+1}
j ← j + 1
410
no
j > n ?
412
yes
414
sum > th3 ?
yes
240
デゾルブ検出処理
no

【図８】

図８



フロントページの続き

(72)発明者　谷口　勝美
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内